



OIPF

ENTERED

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/981,900B

DATE: 03/28/2002 P-5  
TIME: 16:29:17

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF3\03282002\I981900B.raw

3 <110> APPLICANT: Sticklen, Masomeh B  
 4 Maqbool, Shahina B  
 5 Dale, Bruce E  
 7 <120> TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING LIGNINASE AND CELLULASE WHICH  
 DEGRADE LIGNIN  
 8 AND CELLULOSE TO FERMENTABLE SUGARS  
 10 <130> FILE REFERENCE: MSU 4.1-539  
 12 <140> CURRENT APPLICATION NUMBER: US/09/981,900B  
 C--> 13 <141> CURRENT FILING DATE: 2002-03-18  
 15 <150> PRIOR APPLICATION NUMBER: 60/242,408  
 16 <151> PRIOR FILING DATE: 2000-10-20  
 18 <160> NUMBER OF SEQ ID NOS: 22  
 20 <170> SOFTWARE: PatentIn version 3.1  
 22 <210> SEQ ID NO: 1  
 23 <211> LENGTH: 1110  
 24 <212> TYPE: DNA  
 25 <213> ORGANISM: Oryza sativa  
 27 <400> SEQUENCE: 1

28	gggtcggaga	tgccaccacg	gccacaaccc	acgagcccgg	cgcgacacca	ccgcgcgcgt	60
30	tgagccagcc	acaaacgccc	gcgcataggg	gcgcgcgacg	cggccaatcc	taccacatcc	120
32	ccggcctccg	cggtctgagc	gccgtgccat	ccgatccgct	gagttttggc	tatttatacg	180
34	taccgcggga	gcctgtgtgc	agagagtgcg	tctcaagaag	tactcgagca	aagaaggaga	240
36	gagcttggtg	agctgcagag	atggccccct	ccgtgatggc	gtcgtcggcc	accaccgtcg	300
38	ctcccttcca	gggtcgaagt	ccaccgcccg	catgccgtcg	cccgcgcgtc	gaactccagc	360
40	ttcggcaacg	tcagcatggc	ggcaggatca	ggtgcatgca	ggtaattacc	tactgatcca	420
42	acacacattc	ttcttcttct	ttcttcttct	aaccaacatt	aaccaacaac	tcaattatcg	480
44	tttattcatt	gaggtgtggc	cgattgaggg	catcaagaag	ttcgagaccc	tctcctacct	540
46	gccaccgctc	accgtggagg	acctcctgaa	gcagatcgag	tacctagctc	cgttccaagt	600
48	ggtgccctgc	ctcgagttca	gcaaggctcg	atttgtctac	cgtagagaacc	acaagtcctc	660
50	tggtacttac	gacggcaggt	actggaccat	gtggaagctg	cccatgttcg	ggtgcaccga	720
52	cgccaccag	gtcgtcaagg	agctcgagga	ggccaagaag	gcgtaccctg	atgcattcgt	780
54	ccgtatcatt	ggcttcgaca	acgttaggca	ggtgcagctc	atcagcttca	tcgcctacaa	840
56	cccgggctgc	gaggagtctg	gtggcaacta	agccgtcatt	gtcatatata	gcctcgttta	900
58	attgttcatt	tctgattcga	tgatgtctcc	caccttggtt	cgtgtgttcc	cagtttggtt	960
60	catcgtcttt	tgattttacc	ggcgcgtgct	tgcttttggt	ttttcttttc	acctgattct	1020
62	ctctctgact	tgatgtaaga	gtggtatctg	ctacgactat	atgttggttg	ggtgaggcat	1080
64	atgtgaatga	aatctatgaa	agctccggct				1110

67 <210> SEQ ID NO: 2  
 68 <211> LENGTH: 38  
 69 <212> TYPE: PRT  
 70 <213> ORGANISM: Oryza sativa  
 72 <400> SEQUENCE: 2

74	Met	Ala	Pro	Ser	Val	Met	Ala	Ser	Ser	Ala	Thr	Thr	Val	Ala	Pro	Phe
75	1					5				10				15		

## RAW SEQUENCE LISTING

DATE: 03/28/2002

PATENT APPLICATION: US/09/981,900B

TIME: 16:29:17

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF3\03282002\I981900B.raw

```

78 Gln Gly Ser Ser Pro Pro Pro Ala Cys Arg Arg Pro Pro Ser Glu Leu
79                20                25                30
82 Gln Leu Arg Gln Arg Gln
83                35
86 <210> SEQ ID NO: 3
87 <211> LENGTH: 6
88 <212> TYPE: PRT
89 <213> ORGANISM: Artificial Sequence
91 <220> FEATURE:
92 <223> OTHER INFORMATION: Signal peptide targets the peroxisomes of plants
94 <400> SEQUENCE: 3
96 Arg Ala Val Ala Arg Leu
97 1                5
100 <210> SEQ ID NO: 4
101 <211> LENGTH: 3004
102 <212> TYPE: DNA
103 <213> ORGANISM: Acidothermus cellulolyticus
105 <220> FEATURE:
106 <221> NAME/KEY: CDS
107 <222> LOCATION: (824)..(2512)
108 <223> OTHER INFORMATION: E I beta-1,4-endoglucanase precursor
111 <400> SEQUENCE: 4
112 ggatccacgt tgtacaaggt cacctgtccg tcgttctggt agagcggcgg gatggtcacc      60
114 cgcacgatct ctcttttgtt gatgtcgacg gtcacgtggt tacggtttgc ctccggccgcg      120
116 attttcgcgc tcgggcttgc tccggctgtc ggggttcggtt tggcgtggtg tgcggagcac      180
118 gccgagggcga tcccaatgag ggcaagggca agagcggagc cgatggcacg tcgggtggcc      240
120 gatgggggtac gccgatgggg cgtggcgctc ccgccgcgga cagaaccgga tgcggaatag      300
122 gtcacgggtgc gacatgttgc cgtaccgcgg acccggtatg caaggggtggg tgcgcgggtc      360
124 gcctgtgagc tgccggctgg cgtctggatc atgggaacga tcccaccatt ccccgcaatc      420
126 gacgcgatcg ggagcagggc ggcgcgagcc ggaccgtgtg gtcgagccgg acgattcgcc      480
128 catacgggtgc tgcaatgccc agcgccatgt tgtcaatccg ccaaatgcag caatgcacac      540
130 atggacaggg attgtgactc tgagtaatga ttggattgcc ttcttgccgc ctacgcgtta      600
132 cgcagagtag gcgactgtat gcggtagggt ggcgctccag ccgtgggctg gacatgcctg      660
134 ctgcgaactc ttgacacgtc tggttgaacg cgcaatactc ccaacaccga tgggatcggt      720
136 cccataagtt tccgtctcac aacagaatcg gtgcgccctc atgatcaacg tgaaaggagt      780
138 acggggggaga acagacgggg gagaaccaa cgggggattg gcg gtg ccg cgc gca      835
139                                     Val Pro Arg Ala
140                                     1
142 ttg cgg cga gtg cct ggc tcg cgg gtg atg ctg cgg gtc ggc gtc gtc      883
143 Leu Arg Arg Val Pro Gly Ser Arg Val Met Leu Arg Val Gly Val Val
144 5                10                15                20
146 gtc gcg gtg ctg gca ttg gtt gcc gca ctc gcc aac cta gcc gtg ccg      931
147 Val Ala Val Leu Ala Leu Val Ala Ala Leu Ala Asn Leu Ala Val Pro
148                25                30                35
150 cgg ccg gct cgc gcc gcg ggc ggc ggc tat tgg cac acg agc ggc cgg      979
151 Arg Pro Ala Arg Ala Ala Gly Gly Gly Tyr Trp His Thr Ser Gly Arg
152                40                45                50
154 gag atc ctg gac gcg aac aac gtg ccg gta ccg atc gcc ggc atc aac      1027
155 Glu Ile Leu Asp Ala Asn Asn Val Pro Val Arg Ile Ala Gly Ile Asn

```

## RAW SEQUENCE LISTING

DATE: 03/28/2002

PATENT APPLICATION: US/09/981,900B

TIME: 16:29:17

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF3\03282002\I981900B.raw

156	55	60	65	
158	tgg ttt ggg ttc gaa acc tgc aat tac gtc gtg cac ggt ctc tgg tca	1075		
159	Trp Phe Gly Phe Glu Thr Cys Asn Tyr Val Val His Gly Leu Trp Ser			
160	70 75 80			
162	cgc gac tac cgc agc atg ctc gac cag ata aag tcg ctc ggc tac aac	1123		
163	Arg Asp Tyr Arg Ser Met Leu Asp Gln Ile Lys Ser Leu Gly Tyr Asn			
164	85 90 95 100			
166	aca atc cgg ctg ccg tac tct gac gac att ctc aag ccg ggc acc atg	1171		
167	Thr Ile Arg Leu Pro Tyr Ser Asp Asp Ile Leu Lys Pro Gly Thr Met			
168	105 110 115			
170	ccg aac agc atc aat ttt tac cag atg aat cag gac ctg cag ggt ctg	1219		
171	Pro Asn Ser Ile Asn Phe Tyr Gln Met Asn Gln Asp Leu Gln Gly Leu			
172	120 125 130			
174	acg tcc ttg cag gtc atg gac aaa atc gtc gcg tac gcc ggt cag atc	1267		
175	Thr Ser Leu Gln Val Met Asp Lys Ile Val Ala Tyr Ala Gly Gln Ile			
176	135 140 145			
178	ggc ctg cgc atc att ctt gac cgc cac cga ccg gat tgc agc ggg cag	1315		
179	Gly Leu Arg Ile Ile Leu Asp Arg His Arg Pro Asp Cys Ser Gly Gln			
180	150 155 160			
182	tcg gcg ctg tgg tac acg agc agc gtc tcg gag gct acg tgg att tcc	1363		
183	Ser Ala Leu Trp Tyr Thr Ser Ser Val Ser Glu Ala Thr Trp Ile Ser			
184	165 170 175 180			
186	gac ctg caa gcg ctg gcg cag cgc tac aag gga aac ccg acg gtc gtc	1411		
187	Asp Leu Gln Ala Leu Ala Gln Arg Tyr Lys Gly Asn Pro Thr Val Val			
188	185 190 195			
190	ggc ttt gac ttg cac aac gag ccg cat gac ccg gcc tgc tgg ggc tgc	1459		
191	Gly Phe Asp Leu His Asn Glu Pro His Asp Pro Ala Cys Trp Gly Cys			
192	200 205 210			
194	ggc gat ccg agc atc gac tgg cga ttg gcc gcc gag ccg gcc gga aac	1507		
195	Gly Asp Pro Ser Ile Asp Trp Arg Leu Ala Ala Glu Arg Ala Gly Asn			
196	215 220 225			
198	gcc gtg ctc tcg gtg aat ccg aac ctg ctc att ttc gtc gaa ggt gtg	1555		
199	Ala Val Leu Ser Val Asn Pro Asn Leu Leu Ile Phe Val Glu Gly Val			
200	230 235 240			
202	cag agc tac aac gga gac tcc tac tgg tgg ggc ggc aac ctg caa gga	1603		
203	Gln Ser Tyr Asn Gly Asp Ser Tyr Trp Trp Gly Gly Asn Leu Gln Gly			
204	245 250 255 260			
206	gcc ggc cag tac ccg gtc gtg ctg aac gtg ccg aac cgc ctg gtg tac	1651		
207	Ala Gly Gln Tyr Pro Val Val Leu Asn Val Pro Asn Arg Leu Val Tyr			
208	265 270 275			
210	tcg gcg cac gac tac gcg acg agc gtc tac ccg cag acg tgg ttc agc	1699		
211	Ser Ala His Asp Tyr Ala Thr Ser Val Tyr Pro Gln Thr Trp Phe Ser			
212	280 285 290			
214	gat ccg acc ttc ccc aac aac atg ccc ggc atc tgg aac aag aac tgg	1747		
215	Asp Pro Thr Phe Pro Asn Asn Met Pro Gly Ile Trp Asn Lys Asn Trp			
216	295 300 305			
218	gga tac ctc ttc aat cag aac att gca ccg gta tgg ctg ggc gaa ttc	1795		
219	Gly Tyr Leu Phe Asn Gln Asn Ile Ala Pro Val Trp Leu Gly Glu Phe			
220	310 315 320			

## RAW SEQUENCE LISTING

DATE: 03/28/2002

PATENT APPLICATION: US/09/981,900B

TIME: 16:29:17

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF3\03282002\I981900B.raw

222	ggt	acg	aca	ctg	caa	tcc	acg	acc	gac	cag	acg	tgg	ctg	aag	acg	ctc	1843
223	Gly	Thr	Thr	Leu	Gln	Ser	Thr	Thr	Asp	Gln	Thr	Trp	Leu	Lys	Thr	Leu	
224	325					330					335					340	
226	gtc	cag	tac	cta	cgg	ccg	acc	gcg	caa	tac	ggt	gcg	gac	agc	ttc	cag	1891
227	Val	Gln	Tyr	Leu	Arg	Pro	Thr	Ala	Gln	Tyr	Gly	Ala	Asp	Ser	Phe	Gln	
228					345					350					355		
230	tgg	acc	ttc	tgg	tcc	tgg	aac	ccc	gat	tcc	ggc	gac	aca	gga	gga	att	1939
231	Trp	Thr	Phe	Trp	Ser	Trp	Asn	Pro	Asp	Ser	Gly	Asp	Thr	Gly	Gly	Ile	
232				360					365					370			
234	ctc	aag	gat	gac	tgg	cag	acg	gtc	gac	aca	gta	aaa	gac	ggc	tat	ctc	1987
235	Leu	Lys	Asp	Asp	Trp	Gln	Thr	Val	Asp	Thr	Val	Lys	Asp	Gly	Tyr	Leu	
236			375					380					385				
238	gcg	ccg	atc	aag	tcg	tcg	att	ttc	gat	cct	gtc	ggc	gcg	tct	gca	tcg	2035
239	Ala	Pro	Ile	Lys	Ser	Ser	Ile	Phe	Asp	Pro	Val	Gly	Ala	Ser	Ala	Ser	
240		390					395					400					
242	cct	agc	agt	caa	ccg	tcc	ccg	tcg	gtg	tcg	ccg	tct	ccg	tcg	ccg	agc	2083
243	Pro	Ser	Ser	Gln	Pro	Ser	Pro	Ser	Val	Ser	Pro	Ser	Pro	Ser	Pro	Ser	
244	405					410					415					420	
246	ccg	tcg	gcg	agt	cgg	acg	ccg	acg	cct	act	ccg	acg	ccg	aca	gcc	agc	2131
247	Pro	Ser	Ala	Ser	Arg	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Pro	Thr	Ala	Ser	
248				425						430					435		
250	ccg	acg	cca	acg	ctg	acc	cct	act	gct	acg	ccc	acg	ccc	acg	gca	agc	2179
251	Pro	Thr	Pro	Thr	Leu	Thr	Pro	Thr	Ala	Thr	Pro	Thr	Pro	Thr	Ala	Ser	
252				440					445					450			
254	ccg	acg	ccg	tca	ccg	acg	gca	gcc	tcc	gga	gcc	cgc	tgc	acc	gcg	agt	2227
255	Pro	Thr	Pro	Ser	Pro	Thr	Ala	Ala	Ser	Gly	Ala	Arg	Cys	Thr	Ala	Ser	
256			455					460					465				
258	tac	cag	gtc	aac	agc	gat	tgg	ggc	aat	ggc	ttc	acg	gta	acg	gtg	gcc	2275
259	Tyr	Gln	Val	Asn	Ser	Asp	Trp	Gly	Asn	Gly	Phe	Thr	Val	Thr	Val	Ala	
260		470				475					480						
262	gtg	aca	aat	tcc	gga	tcc	gtc	gcg	acc	aag	aca	tgg	acg	gtc	agt	tgg	2323
263	Val	Thr	Asn	Ser	Gly	Ser	Val	Ala	Thr	Lys	Thr	Trp	Thr	Val	Ser	Trp	
264	485					490					495					500	
266	aca	ttc	ggc	gga	aat	cag	acg	att	acc	aat	tcg	tgg	aat	gca	gcg	gtc	2371
267	Thr	Phe	Gly	Gly	Asn	Gln	Thr	Ile	Thr	Asn	Ser	Trp	Asn	Ala	Ala	Val	
268				505						510					515		
270	acg	cag	aac	ggt	cag	tcg	gta	acg	gct	cgg	aat	atg	agt	tat	aac	aac	2419
271	Thr	Gln	Asn	Gly	Gln	Ser	Val	Thr	Ala	Arg	Asn	Met	Ser	Tyr	Asn	Asn	
272			520						525					530			
274	gtg	att	cag	cct	ggt	cag	aac	acc	acg	ttc	gga	ttc	cag	gcg	agc	tat	2467
275	Val	Ile	Gln	Pro	Gly	Gln	Asn	Thr	Thr	Phe	Gly	Phe	Gln	Ala	Ser	Tyr	
276			535					540					545				
278	acc	gga	agc	aac	gcg	gca	ccg	aca	gtc	gcc	tgc	gca	gca	agt	taa		2512
279	Thr	Gly	Ser	Asn	Ala	Ala	Pro	Thr	Val	Ala	Cys	Ala	Ala	Ser			
280		550					555						560				
282	tacgtcgggg	agccgacggg	agggtccgga	ccgtcgggttc	cccggcttcc	acctatggag											2572
284	cgaacccaac	aatccggacg	gaactgcagg	taccagagag	gaacgacacg	aatgcccgcc											2632
286	atctcaaaac	ggctgcgagc	cggcgtcctc	gccggggcgg	tgagcatcgc	agcctccatc											2692
288	gtgccgctgg	cgatgcagca	tcctgccatc	gccgcgacgc	acgtcgacaa	tcctatgcg											2752

## RAW SEQUENCE LISTING

DATE: 03/28/2002

PATENT APPLICATION: US/09/981,900B

TIME: 16:29:17

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF3\03282002\I981900B.raw

```

290 ggagcgacct tcttcgtcaa cccgtactgg gcgcaagaag tacagagcga acggcggaacc 2812
292 agaccaatgc cactctcgca gcgaaaatgc gcgtcgtttc cacatatctg acggccgtct 2872
294 ggatggaccg catcgctgcg atcaacggcg tcaacggcgg acccggttg acgacatatc 2932
296 tggacgccgc cctctcccag cagcagggaa ccacccctga agtcattgag attgtcatct 2992
298 acgatctgcc gg 3004
301 <210> SEQ ID NO: 5
302 <211> LENGTH: 562
303 <212> TYPE: PRT
304 <213> ORGANISM: Acidothermus cellulolyticus
306 <400> SEQUENCE: 5
308 Val Pro Arg Ala Leu Arg Arg Val Pro Gly Ser Arg Val Met Leu Arg
309 1 5 10 15
312 Val Gly Val Val Val Ala Val Leu Ala Leu Val Ala Ala Leu Ala Asn
313 20 25 30
316 Leu Ala Val Pro Arg Pro Ala Arg Ala Ala Gly Gly Gly Tyr Trp His
317 35 40 45
320 Thr Ser Gly Arg Glu Ile Leu Asp Ala Asn Asn Val Pro Val Arg Ile
321 50 55 60
324 Ala Gly Ile Asn Trp Phe Gly Phe Glu Thr Cys Asn Tyr Val Val His
325 65 70 75 80
328 Gly Leu Trp Ser Arg Asp Tyr Arg Ser Met Leu Asp Gln Ile Lys Ser
329 85 90 95
332 Leu Gly Tyr Asn Thr Ile Arg Leu Pro Tyr Ser Asp Asp Ile Leu Lys
333 100 105 110
336 Pro Gly Thr Met Pro Asn Ser Ile Asn Phe Tyr Gln Met Asn Gln Asp
337 115 120 125
340 Leu Gln Gly Leu Thr Ser Leu Gln Val Met Asp Lys Ile Val Ala Tyr
341 130 135 140
344 Ala Gly Gln Ile Gly Leu Arg Ile Ile Leu Asp Arg His Arg Pro Asp
345 145 150 155 160
348 Cys Ser Gly Gln Ser Ala Leu Trp Tyr Thr Ser Ser Val Ser Glu Ala
349 165 170 175
352 Thr Trp Ile Ser Asp Leu Gln Ala Leu Ala Gln Arg Tyr Lys Gly Asn
353 180 185 190
356 Pro Thr Val Val Gly Phe Asp Leu His Asn Glu Pro His Asp Pro Ala
357 195 200 205
360 Cys Trp Gly Cys Gly Asp Pro Ser Ile Asp Trp Arg Leu Ala Ala Glu
361 210 215 220
364 Arg Ala Gly Asn Ala Val Leu Ser Val Asn Pro Asn Leu Leu Ile Phe
365 225 230 235 240
368 Val Glu Gly Val Gln Ser Tyr Asn Gly Asp Ser Tyr Trp Trp Gly Gly
369 245 250 255
372 Asn Leu Gln Gly Ala Gly Gln Tyr Pro Val Val Leu Asn Val Pro Asn
373 260 265 270
376 Arg Leu Val Tyr Ser Ala His Asp Tyr Ala Thr Ser Val Tyr Pro Gln
377 275 280 285
380 Thr Trp Phe Ser Asp Pro Thr Phe Pro Asn Asn Met Pro Gly Ile Trp
381 290 295 300
384 Asn Lys Asn Trp Gly Tyr Leu Phe Asn Gln Asn Ile Ala Pro Val Trp

```

Use of n and / or Xaa has been detected in the Sequence Listing. Review the Sequence Listing to ensure a corresponding explanation is present in the <220> to <223> fields of each sequence using n or Xaa.

## VERIFICATION SUMMARY

DATE: 03/28/2002

PATENT APPLICATION: US/09/981,900B

TIME: 16:29:18

Input Set : A:\PTO.VSK.txt

Output Set: N:\CRF3\03282002\I981900B.raw

L:13 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:510 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6  
L:511 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6  
L:518 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6  
L:519 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6  
L:558 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6  
L:559 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6  
L:672 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7  
L:680 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7  
L:720 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7